DOCUMENT RESUME

ED 355 267 TM 019 618

AUTHOR Rudner, Lawrence M.
TITLE The Achievement Testing Provisions of the Virginia Home Schooling Requirements. Testimony for the Virginia State Legislature.
PUB DATE 4 Feb 93
NOTE 8p.
PUB TYPE Legal/Legislative/Regulatory Materials (090) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; *Cutting Scores; Decision Making; Elementary Secondary Education; *Home Schooling; Measurement Techniques; *Minimum Competency Testing; Norm Referenced Tests; Portfolios (Background Materials); *State Legislation; State Programs; Student Evaluation; *Testing Problems; Testing Programs; Test Norms
IDENTIFIERS Percentile Ranks; Stanines; Testimony; *Virginia

ABSTRACT

The requirement that home schooled students score at or above the 40th percentile on a nationally standardized test runs counter to several widely accepted principles of measurement. A single score should not be used alone in decision making, and there is no evidence that this is a valid use of the test. In addition, the 40th percentile is not an appropriate demarcation point for typical, or "normal," performance. A review of the way in which norm-referenced tests are developed confirms that such a requirement is not appropriate. If a set passing score is desired, the fourth stanine is widely accepted demarcation for typical or normal. The use of Virginia or school district norms that are no more than 5 years old is suggested as an alternative to national norms. In addition, superintendents should be urged to consider information in addition to test scores when making a home school judgment. Virginia should encourage superintendents to consider portfolios as an additional assessment option. These suggestions are consistent with widely accepted measurement practices, and will serve better than the language as it stands. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

The Achievement Testing Provisions of the Virginia Home Schooling Requirements

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Lawrence M. Rudner
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

Testimony for the Virginia State Legislature
February 4, 1993

Lawrence M. Rudner, MBA, Ph.D.
ERIC Clearinghouse on Tests, Measurement, and Evaluation
3333 K Street NW, Washington, DC 20007

ED355267

My name is Lawrence Rudner, I have been asked to testify as a measurement expert on the requirement that home schooled students score at or above the 40th percentile on a national recognized standardized test. I will explain how norm referenced tests are developed, implications with regard to this law, what is wrong with the requirement and why, and then make some recommendations.

The requirement runs counter to several widely accepted measurement principals. A single score should not be used alone in decision making. There is no evidence that this is a valid use of a student achievement test and logic indicates that is not a valid use. Home schooling parents that are excellent teachers will have children that fail the test. Home schooling parents that are horrible teachers will be permitted to teach. Further the 40th percentile is not the appropriate demarkation point for typical or "normal" performance.

I will recommend that the 4th stanine be used rather than the 40th percentile, that the norming group be defined, and that superintendents be encouraged to rationally consider other forms of evidence of satisfactory progress.

How norm referenced tests are developed

Norm-referenced tests, such as the Iowa Test of Basic Skills, the Stanford Achievement Test, and the California Achievement Test, help you compare one student's performance with the performances of a large group of students. They are designed to make distinctions between students' performances and pinpoint where a student stands in relation to the normative group -- a large group of students used in test development.

When developers create norm-referenced tests, they carefully survey existing curricula so that they can write test items to reflect the material that has been taught in most schools. Based on this analysis, they prepare detailed test specifications, or test blueprints, that outline the curricular objectives that will be measured and the number of items that will be used to assess each objective. These objectives then guide the developers in writing the test items.

To ensure that the final test has a sufficient number of high-quality items in each curricular area, developers usually pilot test items on a sample of students using two to three times as many items as is planned for the final version of the test. In developing tests that are

TM019818

2 BEST COPY AVAILABLE

going to be used nationally. these "tryout samples" closely match the U.S. student population in terms of such variables as community size, geographic region, family income, years of parental schooling, and nationality. Based on the results of the pilot test, developers retain only the test items that meet certain statistical standards.

To use the Primary II battery of the Stanford Achievement Test as an example, 2,565 items were piloted and 1,326 items were retained for the three final forms.

One of the most important criteria for deciding whether to retain a test item is how well that item contributes to the variability of test scores. Good test developers compose items that encourage variability. They create items that are neither too easy nor too hard and then use the item tryout to confirm their decisions. Items that are too easy or too difficult do not contribute to the variability of test scores and usually will be eliminated.

After developers select the items for a test, they develop test norms and normative test scores, such as grade equivalent scores and percentiles. These norms provide a means to compare the performance of one student or group of students with the performance of a specified reference group. While it is possible to have several reference groups, most standardized achievement test batteries use a representative sample of the U.S. population of school children as the benchmark.

In context, these norms have meaning for most school systems. The norms describe the typical performance of U.S. students on these items at the time the norms were developed.

What are the advantages of norm-referenced tests? They allow you to analyze the general progress of large groups of students. They give you a basis for examining an individual student's general performance at a given point in time.

What are the limitations of norm-referenced tests? They are inappropriate for following an individual student's progress in specific skills. They assess a relatively narrow range of desired educational outcomes. Norms become quickly outdated.

Implications

1. A person's test score does not fully reflect what he or she has mastered.

Numerous relevant skills are not going to be on any test that is used. The original blueprints were designed to reflect a common curriculum, not the curriculum in Virginia. Relevant items were eliminated because they contributed little to the statistical quality of the test. Test items do not always elicit the desired behavior - that's why so many are eliminated in field testing.

2. Test scores contain error

The difference between a person's true ability in the domain covered by a test and their performance on the items on the test which were sampled from that domain. Some of that error is due to errors in test design, some to the inexact art of writing test items, and some are due to the nature of measuring any cognitive function. People are not always consistent.

While percentile scores are usually reported in increments of 100 (38th percentile, 39th percentile, etc), they are typically only accurate to the nearest six one-hundredths (6 percentiles). If a person whose true ability is exactly at the 50th percentile took a test 100 times, then 68 out of those 100 times he would have a score between the 44th and 56th percentile; 95 times out of 100 he would score between the 38th and 62nd percentile. There is no significant difference between test scores that are 12 percentiles apart. Because scores flop around in these intervals, coarser reporting categories, such as stanines, better reflect the accuracy of the test.

3. Percentile scores are clustered in the middle

Percentile scores tell you the percent of students in the norming sample whose scores were at or lower than a given raw score. They are among the most commonly reported scores and are best used to describe a student's standing relative to the norming group at the time of testing. For example, if a student's raw score is in the 80th percentile, then that student scored equal to or higher than 80% of the students who took the test when the test was normed.

The percentile scale is based on the familiar bell curve. The number of people is greatest in the middle and tapers toward the ends. Because so may people have scores near the middle, a small increase in the number right can greatly increase the percentile score. For example, a raw score of 22 items on the Comprehensive Test of Basic Skills Third Grade Reading Vocabulary subtest is at the 48th percentile. If a person misses 1 more item, they would drop to the 41st percentile; missing two more raw items would place the child at the 34th percentile - a drop of 14 percentile points. There is nothing between the 34th and 41st percentile in 3rd grade reading vocabulary. A two item drop for someone at the 94th percentile would drop their percentile score only 3 percentile points.

4. National norms have major limitations

It should also be recognized that the percentile score is based on the norming group. Over time, norms become outdated. Part of the recent headlines about the national average being above average was due to several states using outdated tests. Thus a student can have entirely different percentile scores as a result of shopping around and picking a test with a different match to the curriculum or different norms. Further, there is a question as to which norms to use. Publishers often provide multiple norms for the same test. While the use of a single cut-score appears to be cut-and-dry and therefore an objective safeguard, it is an ill-defined criterion that can be readily be abused.

5. The average student will not always be at the 50th percentile.

The 50th percentile is defined as the mathematical average score of all the students in the large normative group. Half the students are below the 50th percentile, half are above. The norms are redefined for each test and grade level. For each grade level different groups are used. Thus, while the group average will always be the 50th percentile, an average student will change. It is unlikely that a given student who is at the 50th percentile in third grade will be at the 50th percentile in the forth grade. Individuals fluctuate greatly over their educational career. Thus, the average student is not always average.

What is wrong and why

1. A single score should not be used as the sole decision making criterion

While scores appear to be object, hard and fast indicators, they are not. Tests are not that accurate, people are not that consistent. The *Code of Ethics for Test Use*, which was signed by all major test publishers, and the *Standards for Educational and Psychological Tests*, published by several professional organizations and recognized in numerous legal cases, specifically state that a single test score should never be used as the sole criterion in decision making. Rather related factors and other information should be considered when evaluating test scores. The SAT, for example, is never used as the sole criterion in college admissions. The high school record, recommendations, involvement in extracurricular activities are always included in making admissions decisions.

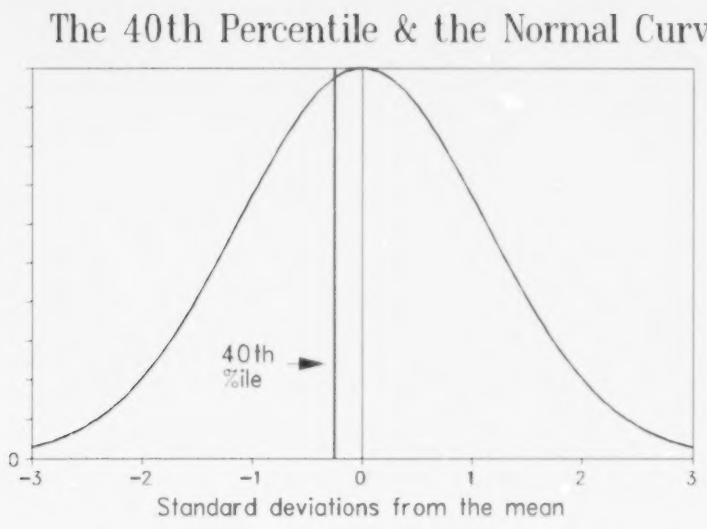
2. The 40th percentile is the wrong cut-score

Because of the clustering of scores, the 40th percentile is extremely close to the 50th percentile. I already pointed out that missing 2 questions can drop a student from the 48th to the 34th percentile. The next figure shows where the 40th percentile appears on the normal curve -- there is hardly any distance between the mean, which is the 50th percentile, and the 40th percentile. To make this point even clearer, I will relate the 40th percentile to scores that people are more familiar with -- the IQ scale and the SAT scale. The 40th percentile corresponds to an IQ of 96 and an SAT Mathematics score of 445 (the 1992 SAT Mathematics mean was 476). These scores are virtually identical to the national averages of 100 for IQ tests, and 476 for the SAT.

There are numerous ways to select a defendable cut-score. Some low cost and logically defendable ways are to: a) use a widely accepted definition of "normal", b) identify the percentile score that typically delineates students who are promoted from grade to grade and those that are not, and c) identify the percentage of students that are typically left back.

3. There is no evidence that this is a valid use of a norm referenced student achievement test

By requiring a student to obtain a certain score in order to remain a home schooler, you are using student achievement tests as licensure examinations -- a purpose that was never



The 40th percentile is virtually identical to the 50th percentile.

intended by the test publisher, and a purpose which is probably not valid. Further you are holding home schoolers to a standard well above that which you expect of licensed public school teachers.

The state has every right to expect teachers, be they home schoolers or public school teachers, to be competent. How do you measure this competency? Public schools use adequacy of lesson plans, classroom observation, peer evaluations, student ratings, and test scores of large groups of students as indicators. They typically consider all these indicators in evaluating a teacher. It is ludicrous to think that a teacher would be fired for poor performance on only one of these measures. It is even more ludicrous to think that a teacher would be fired because one student once did poorly on a test.

The state also has a right to expect students, be they home schoolers or in any other form of schooling, to make satisfactory progress. How do you measure progress? You use pre-test and post test scores. In the case of home schooling, you can look at unit mastery tests. If you prefer to look at test scores on a nationally standardized test, you can examine a child's relative standing at the end of one year in comparison to his standing at the end of the next year.

There is still a problem with using test scores by themselves as indicators of teaching quality and progress. Some students have behavioral problems, others have low ability levels. As a result, children learn at different paces and master different skills at different times.

4. The norming group has not been defined.

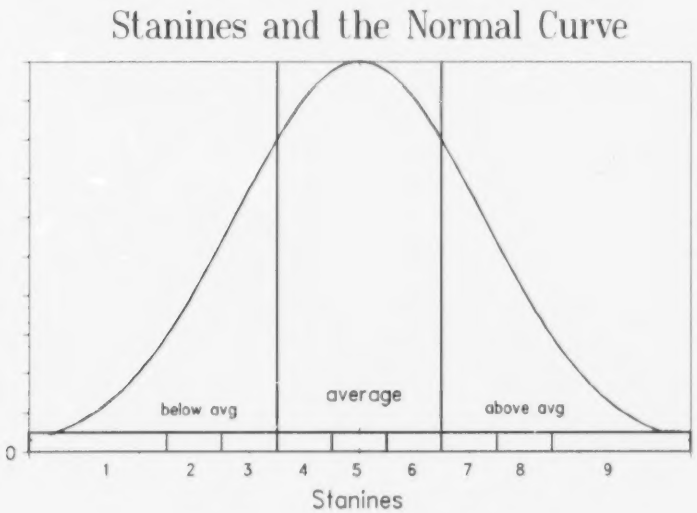
Key to interpreting norm referenced achievement tests is the norming group that was used in test development. The law does not specify either the age of the norms or its

composition. The current trend is for outdated norms to result in higher scores than current norms. For example, the 40th percentile on the current Comprehensive Test of Basic Skills (CTBS/4) normed in 1988 and published in 1990 is equivalent to the 50th percentile on the previous CTBS (CTBS/U) normed in 1981. Further, since norms are available for the nation, state, and district, the law needs to clarify the comparison basis.

Recommendations

I fully recognize the desirability of having a pre-specified passing score. A known criteria can reduce subjectivity and therefore greatly increase the integrity and effectiveness of the decision making process. It can minimize conflict between the home schooling parent and the public school superintendent.

If a set passing score is desired, then the forth stanine is the appropriate and widely accepted demarkation for "typical" or "normal". Stanine is short for *standard nine*. The stanine scale is formed by dividing the total score distribution into nine units. Except for the first and last unit (the tapered tails), each of the units is 1/2 of a standard deviation wide. Stanines are used as an alternative to percentiles, because they better reflect the accuracy of standardized norm referenced achievement tests. The next figure relates stanines to the normal curve.



Scores in stanines 4, 5 and 6 are the usual and customary definition for "average" or normal performance. This corresponds to -.75 to .75 standard deviations from the mean and IQ scores of 90-110.

The 4, 5, and 6th stanines are widely accepted as "average". You will find such statements in virtually every test manual and every measurement textbook. Approximately

54% of the population will be expected to have scores in that range. On the IQ scale, this corresponds to IQs of about 90-110. In terms of standard deviation units, this corresponds to -.75 to +.75 standard deviations.

An issue still remains with regard to the norming or reference group. Rather than national norms, I would suggest using Virginia or district norms that are not more than 5 years old. Rather than stating that a student must be of at least average performance in comparison to the nation, this would require that the home schooler be of at least average performance with regard to other students in the region. It is quite conceivable that a student could be slightly below average for the nation, yet well above average for the district. Requiring that such a child attend the district school would not necessarily be in his or her best interests. To protect home schoolers in affluent neighborhoods, I would require home schoolers to be at or above the 4th stanine using either state or district norms.

Finally, superintendents should be encouraged to consider information in addition to test scores when making a home school judgement. For example, shouldn't a learning disabled child that was doing horribly in school and then made steady gains in a home schooling situation be permitted to stay in the home school? Shouldn't a child that does horribly on standardized tests but shows consistently shows average or above average ability on curricular tasks be permitted to stay in a home school. Shouldn't superintendents consider the quality of lesson plans and progress toward curricular objectives in addition to test scores?

The state of West Virginia did not have the foresight of Virginia and did not include a portfolio option in their law. As a result a student who went from the 17th percentile to the 38th percentile was deemed not to have made satisfactory progress and the home schooling rights of the parent were terminated. Virginia should encourage superintendents to exercise the portfolio option and rationally consider other forms of evidence that a child is benefiting from home schooling.

Summary

I have pointed out that what appears to be simple and straight forward language in the law is neither. The 40th percentile requirement is vague and, from a measurement point of view, inappropriate. It fails to recognize the considerable coarseness of standardized norm referenced tests and it fails to define the reference group. I have offered you alternative approaches that are consistent with widely accepted measurement principles.

References

American Psychological Association (1986) *Standards for Educational and Psychological Tests and Manuals*. Third edition. Washington, DC: author

National Council on Measurement in Education (1989) *Code of Ethics for Test Use*. Washington, DC: author.